情報検索結果の知的提示のための自動要約ならびにインタフェースに関する研究（研究課題番号：16016236)

研究代表者　森 辰則　　横浜国立大学大学院・環境情報研究院・助教授
研究分担者　田村 直良　横浜国立大学大学院・環境情報研究院・教授

1　研究の概要

近年，検索エンジンのような情報検索システムが広く利用されるようになり，検索要求に関連のある文書を容易に得る事が出来るようになった．しかし，検索要求に関連性の低い文書を完全に排除できない，検索結果文書の構造化がなされていないなどの問題点がある．有効な方策としては，検索された各文書の要約の提示することによる元文書参照時間の削減，検索結果文書をクラスタリングすることによる関連文書と不要文書の分類などが従来提案されている．これらの方策に共通する必要不可欠な技術は検索結果文書集合を考慮した重要語の抽出である．その代表手法は検索要求中の語の重要度を高くする方法であり，これを自動要約に利用したものが Query-biased Summarization[TS98] である．これらの手法は直観的ではあるが，次の問題点がある: a) 検索エンジンによる各種フィードバック等の工夫が反映されない，b) 利用者の情報要求がトピック等に対応する一つの検索要求で表現できるという仮定に基づいているが，一般にはトピック以外にも個別の事項について詳しく知りたいといった要求があるのが普通である，c) 語を単位として処理を行なうため，文書に記述される語と語の関係を考慮した要約は行なえない．

そこで，本研究課題では上記問題について以下の観点から取り組んだ．研究全体の構成を図 1 に示す．

1. 検索質問だけではなく検索文書間の関係から語の重要度を計算する手法を構築した．これは，複数文書間の類似性構造というマクロな情報を，語の重要度というミクロな情報に写像する新しい手法であり，文書分類の過程で得られた文書間の類似性構造を適切に説明する度合を語の重要度とできる．

2. 利用者の持つ複数の情報要求に対してその答と関連情報を一度に外観できる要約文書の生成を行なうために，質問応答エンジンを文重要度計算に利用する手法を検討した．この質問応答エンジンは事実に関する単独の質問に回答するシステムであるが，その計算過程で得られる，解としての良さを表す評価値を重要度として扱っている．

3. より粒度の細かい複数文書要約では，個々の文書の持つ情報構造を同一の枠組で捉える必要があるため，特定の領域に依存しつつも精度良く文書からそのスキーマを抽出する手法を提案・評価した．

2　研究期間と研究経費の配分額

本研究は，平成 13 年度より 17 年度までの 5 年間に亙って進められた．この間に配分された研究経費を下表に示す．

年度 (平成)	13	14	15	16	17	合計
研究費 (千円)	5,500	5,300	5,100	5,100	5,900	26,900

3　研究成果

3.1　文書分類に基づく情報ナビゲーションと文書要約に関する研究

本部分課題の目的は，情報検索の結果として得られた文書集合から，ユーザーの必要とする文書を効率よく選択するための情報ナビゲーションを提供することである．先行研究である Scatter/Gather 法 (以下，S/G 法)[CKT92] は検索結果文書をクラスタリングし，構造化することによりナビゲーションを行なうが，分類結果を如何にして利用者に説明するかは議論の余地がある．一方で，我々は検索文書間の関係から重要語を抽出するために，文書クラスタリングにより得られた複数文書間の類似性構造というマクロな情報を，

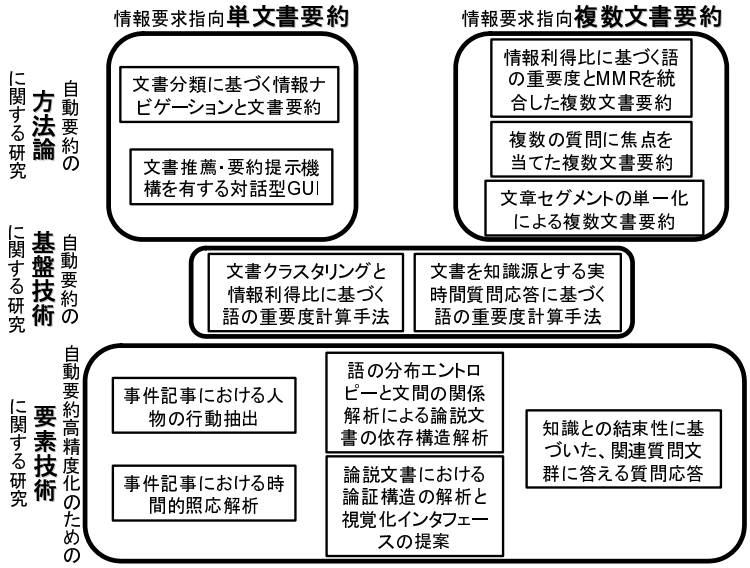

図 1: 研究の概要

語の重要度というミクロな情報に写像する新しい手法を提案している．そこで，本部分課題では，この重要語抽出手法が方法論として文書分類と自然に融合している点に着目し，文書分類に基づく情報ナビゲーションと文書要約を同時に兼ね備えた情報ナビゲーションシステムを提案・考察を行なった．

S/G 法では，検索結果文書を決められた数のクラスタに分類し (Scatter)，各クラスタに対し，それを説明するキーワードを付与する．利用者はその説明記述を基に，必要な文書集合を選択する (Gather)．これらの処理を繰り返す事により文書集合が絞り込まれていく．ここで，利用者インタフェースとして重要な，各クラスタの説明記述については高頻度語を提示するという非常に素朴な戦略が採用されていた．

さて，上記説明記述の要件としては，以下の二点が考えられる: a) クラスタを代表するものであること，b) 他のクラスタとの差異を明らかにするものであること．クラスタ内の高頻度語に注目する手法は，上記第一点目に注目したものである．しかし，この方法では他クラスタの単語情報は一切考慮されず，差異に関する情報は考慮されないため，クラスタ選択の際に重要となる他のクラスタとの違異点が明確に示されない．そこで部分課題では他のクラスタとの差異を測る指標として情報利得比 (informaton gain ratio, IGR) を採用し，この情報利得比を用いた単語重要度を従来手法である TFIDF 法と組み合わせることによりクラスタへの説明記述の生成に利用することを検討した．この手法では，利用者が S/G の過程においてどのクラスタを選択したかという情報，すなわち，ナビゲーションの動的過程の情報が単語重要度の計算に反映される点が興味深い．さらに，ナビゲーションの任意の時点において利用者が指定した文書を上記単語重要度に基づいて要約し，その結果を提示する手法も提案した．

本提案手法を実装したシステムを用いて特定の情報を探索するタスクを数名の被験者に行なってもらう実験を定量的評価として行なった結果，情報要求に対する関連情報 (答) に到達するまでの平均ステップ数による評価によれば，Scatter/Gather で用いられていた単純な TF 手法より，情報利得比を利用する手法の方が優れているという結果が得られた．

3.2 文書推薦・要約提示機構を有する対話型 GUI に関する研究

節 3.1 に示した部分課題の成果を受けて，Scatter/Gather 法に基づくナビゲーション過程において，利用者によるクラスタの選択過程から適合性フィードバックにより利用者の情報要求を自動的に取得・再検索を行なうことにより，新しい文書集合を推薦する機構を考察した．このとき，新文書集合を独立した窓にクラスタリングして提示することにより，利用者がすでに行なっている文書の絞り込み過程は保持しつつ，新しい文書を推薦することを可能とした．利用者は必要に応じて双方のクラスタ集合から自分の情報要求に則したクラスタを選択することができるので，文書推薦と Scatter/Gather の過程を自然に融合している．さらに，検索結果の文書ならびにそのクラスタ構造を Hyperbolic Tree 等を用いて視覚的に提示するグラフィカルユーザインタフェースを構築し，より直観的な閲覧を支援することを可能とした．

3.3 情報利得比に基づく語の重要度と MMR を統合した複数文書要約手法に関する研究

途中の文書クラスタを概観する要約を生成することを目的として，ナビゲーション過程や結果に現れる複数文書を対象とし，複数文書要約を生成する際の基本手法を検討した．上記目的に適する要約文書は「原文の主要内容の網羅性の高さ」ならびに「可読性の高さ」という二つの尺度において高い評価を受けるものである．先行研究としては，可読性を考慮しつつ，内容の網羅性に中心を据えて議論を行なっているものが多い．それらの手法は類似した文書グループの発見において，クラスタリングを用いているが，その一方で，クラスタ間の関係や，各クラスタ内部にある部分クラスタの構造は利用していない．一方，我々は，対象文書集合のより詳細な類似性構造を階層的なクラスタリングにより得ることで，各クラスタでの共通事項のみならず，クラスタ間の差異も考慮できると考えた．そして，この情報を要約の手がかりとすることができれば，より肌理の細かい重要箇所の抽出ができるのではないかと考え，次のように近接した．

まず，節 3.1 に述べた手法により文書集合の間に存在する類似性構造を階層的クラスタリングによりクラスタ構造として抽出し，これを基に語の重要度を計算する．この手法ではクラスタ間に存在する構造，すなわち，差異や共通点を同時に考慮して語の重要度重みに反映できる．これにより，各文書における重要文が検出できる．さらに，複数文書から単一の要約文書を生成する場合には，文書間での内容の重複があり得るのでこれを考慮する必要がある．そこで，我々は，元来，検索質問文とパッセージ群の関連度とパッセージ間の冗長性を同時に扱う要約手法として提案されている Maximal Merginal Relevance(MMR)[CG98] を，単文の重要度と冗長性を同時に考慮する手法として改訂した．

上記手法について，評価型ワークショップ NTCIR3 TSC2 課題 B に参加し，評価をおこなった．その結果，同手法は先行研究である Stein 法などと比較して，原文書集合に対する内容の網羅性において優れているという結果がえられた．特に,「要約率が小さい時」「対象とする文書集合の数が少ない時」という状況下においてその傾向が顕著に表れた．

一方で，文書数が増加するに従って，文間の繋がりの良さ (結束性) が次第に悪くなる傾向にあった． そこで，抽出対象となる重要文の間の結束性を高めるために，文重要度の平滑化を行なう手法を検討した．具体的にはハニング窓関数を用いて各文の重要度を周囲のある範囲の文の重要度も考慮しつつ再計算する．同じく NTCIR3 TSC2 課題 B の設定において，数名の被験者による要約文書の順位づけ評価を行なった．その結果，ハニング窓関数を組み込み，文間の結束性を向上させる事で，要約率が大きい時の内容の網羅性と，全トピックに亙っての可読性に対して，要約の品質が改善されることが示された．

3.4 自動文書要約と質問応答システムの関連に関する研究

3.4.1 複数の質問に焦点を当てた複数文書要約手法

近年,「質問の答に焦点を当てた要約」が注目されている [HSI01]．これは，情報検索過程においては利用者が情報要求を持っており，また，それらが質問文として記述できるという考え方に基づく．複数文書要約においては内容把握ができるように，ある程度の要約文書量が必要であるので，利用者の知りたい事柄の一つ一つについて別々の要約文書を生成すると，最終的に利用者が読むべき文書量が増えてしまう．複数の要求の答とその背景知識を一度に概観できるような要約が生成できることが望ましい．

以上を踏まえて，本部分課題では，複数の質問文に対応可能な文重要度計算法として質問応答エンジンの解のスコアを利用する手法を提案した．ここでは，要約対象文書群は，情報検索等の結果として得られており，また，利用者の情報要求は，複数の質問文として与えられているとする．質問文については，利用者がシステムとの対話の中で一つずつ与えていき，その都度，その答を含む文脈をそれ以前の要約文書との関連を考慮しつつ要約していくという設定が自然ではある．しかし，第一次近似として，複数の質問文が同時に与えられることを想定し，利用者との対話の中での要約生成は今後の課題としたい．

この状況下では，複数文書要約のために，1)「情報要求を考慮した重要箇所抽出」，2)「文書間の冗長箇所の削除」，3)「文書間の相違点の抽出」が必要であると考える．提案手法では，これらについて，a) 質問応答エンジンの出力スコアに基づく文の重要度計算 (上記 1 に対応)，b) 語の出現分布に関する情報利得比に基づく文の重要度計算 (上記 1，3 に対応)，c)MMR に基づく要約文書中の冗長性の制御 (上記 2 に対応) を用いる．更に抽出文間の結束性の担保のために，d) ハニング窓関数に基づく文重要度平滑化を採用した．

与えられた質問文集合を扱うために，我々は独自に開発をしている質問応答システムを用いる．このシステムは，対象文書に対する前処理を必要としない仕組みとなっているので，利用者は簡単なラッパープログラムを記述するだけで，様々な検索エンジンを利用できるという特徴を持つ．このシステムでは，質問文が与えられた後に，形態素解析や構文解析，固有表現抽出などといった計算コストの大きい処理をするので，A^* に基づく解の探索制御ならびにより少ない処理コストで計算のできるスコアの近似手法を導入している．これらにより，文書中の無関係な箇所の計算コストを削減し，実時間処理が行なえる．

同システムのエンジンは質問文が 1 つ与えられると対象文書中の各語 (形態素) に対して解としての適切

さを表すスコアを付与する． 我々はこのスコアを質問の解に注目した時の語の重要度と考え，文の重要度をそれらから計算することを提案する．このスコアを利用することにより，質問に含まれる語や質問型の情報のみを利用する従来手法よりも精度の高い要約生成を行なえると期待される．

評価型ワークショップ NTCIR4 TSC3 における Formal Run の課題により提案手法に基づくシステムを評価した．1) モデル抜粋との比較による抜粋の性能，ならびに，2) モデル要約との比較による質問に対する解の被覆率に基づき評価を行なった結果，提案手法は各種ベースラインよりも高い性能を有することが示された．

3.4.2 知識との結束性に基づいた関連質問質問文群に答える質問応答に関する研究

この部分課題では，前節 3.4.1 の手法を改善すべく，一連の関連質問文群 (以下，文脈依存型質問群) に答える質問応答に関する研究を行なった．前節の手法では複数の質問に焦点を当てた要約文書を生成可能であったが，各質問の間の関連は想定されていないため，各質問に必要十分な文脈が記されていることが前提であった．ところが利用者が互いに関連のある複数の質問を行なう場合には，通常，先の質問を文脈としてそれ以降の質問がなされるため，後の質問では省略や照応表現による参照が生じることが自然である．これに対応するためには，文脈依存型質問群に答える質問応答の技術が必要となる．

本部分課題では，既存の非文脈依存型質問応答を用いて，文脈依存型質問群に回答する一手法を提案した．文脈依存型質問文は一般に代名詞，省略，照応といった参照表現を持つのが普通であるが，もしも，それら参照表現が適切に先行詞に置き換えられれば一問一答型質問応答システムにおいても適切に答を求めることができると期待される．談話理解の分野においては，センタリング理論のように「文脈との結束性に基づく」照応解析手法が多数提案されている．この種の手法は，新しく発話された文とそれ以前の文脈との間の結束性 (結び付き) を最大化するように最適な参照の解釈を見つけ，大抵の場合，適切に働く．しかしながら，省略の検出における曖昧性解消は行なえない．

本部分課題において，我々はもう一方の極端にある照応解析手法である「知識との結束性に基づく手法」，ならびに，それを用いて文脈依存型質問文群に答える手法を提案した．知識との結束性に基づく手法とは，新しく発話された文と知識源と間の結束性を最大化するようにその新しい文の解釈を見つける方法である．質問文の中の参照表現の先行詞同定においては，先行文脈に登場した表現を先行詞候補とし，参照表現を先行詞で補完した質問文と知識源との間の結束性が高くなるように先行詞を選択する．「知識との結束性に基づく手法」は現在のところ次のように実装されている: 1) 与えられたある質問文について，それが持つ各参照表現を可能な先行詞候補の各々で補完することにより，可能性のある (補完された) 質問文候補を生成する．この補完において，I) 意味的な制約に基づく最低限の選択制限のみを考慮する戦略と II) それに加えてセンタリング理論を考慮する戦略が考えられる．2) 各補完質問文候補について「知識との結束性の度合」を計算する．ここで我々はその度合は補完質問文から得られた解の良さ，例えば，一問一答型質問応答システムによって得られた解のスコアに比例していると考えている．3) 最良のスコアを与える補完質問文候補とその答を求め，元の質問文に対する最良の解釈ならびに解とする．

評価型ワークショップ NTCIR-5 QAC3 における評価によれば，戦略 I は質問文群における焦点が絞られている場合 (情報収集型) のほうが焦点が移り変わっていく場合 (ブラウジング型) に比べて得意であることが分かった．一方，戦略 II はブランジング型も適切に処理が可能であり，バランスが良いことが分かった．

3.5 文章セグメントの単一化による複数文章要約

新聞記事やニュース記事のような出来事について書かれた文章に注目してみると，一つの出来事に関する記述が複数の文章にまたがることが多い．また，そのような場合には単に出来事の続きが記述されているだけでなく，視点を変えた文章なども存在する．さらに，複数の文書にわたった新聞記事は，内容が重複する部分が存在する．この研究では，文章をまとめる観点として時間を用いて複数文書自動要約を実現する手法を提案した．文章中に出現する時間表現を手がかりに，事件展開構造を生成する．ここから，様々な観点で事件展開のパスを選択する事によって，多様な要約文を生成することができる．

本研究で提案する事件展開構造とは，文をノードとし，文が表す事象の時間的な前後関係をリンクとして文章を構造化したものである．ノードの時間情報に前後関係があるものは時系列順に整列し，時間情報が同一，もしくは前後関係が不確かなものは並列に並べることにより構造を生成する．この構造から，リンクをたどりながらノードを選択していくことで，時間に沿った文章を生成することができる．

要約文を生成する時，どのような条件でパスを選ぶかを決定する必要がある．ここで用いる条件により，生成される要約文が変化する．これは，条件の選択と同時に要約文の観点を選択しているためである．

この研究では実験データとして，1993 年毎日新聞のコーパスから 2 文書にわたる 2 つの事件について人手で事件展開構造を生成し，パス選択を行うことによって要約文を生成した．

3.6 新聞事件記事の人物の行動抽出

3.6.1 事件記事における人物の行動抽出

この研究では，ある程度実用規模での文書理解，情報抽出を前提とし，文章要約や二次利用可能な情報蓄積を利用目的と想定し，意味構造を検討する．実際には,「犯罪」,「事件」について書かれた新聞記事 (事件記事) を対象とし,「犯罪」,「事件」の意味を表現する「犯罪スキーマ」を提案した．

事件記事に対し，文章に内在する既存の意味関係の解析 (時間セグメント，主題構造解析，語彙連鎖構造解析，文末構造解析) を行い文章を構造化し汎用的内部表現を得る．そして，得られた汎用的内部表現から犯罪スキーマの抽出を行う．犯罪スキーマでは，主に表層文とのパターンマッチング，表層的フレームの解析を行うことにより，記事から，犯罪に関わる動機，犯罪のタイプ，人物，その役割，そのプロフィール，その行動などを抽出する．ある種の情報の抽出には，深い解析を行うよりも，むしろ，表層的な手がかり表現やパターンマッチングを用いることにより，容易に行うことができるものもある．一方，深い解析が必要な情報抽出もある．この解析システムでは，抽出する要素の性質に応じて，両者を組み合わせている．

すべての事件記事は，罪状，動機，供述，人物の 4 つの要素で表現できると仮定する．そこで，我々は事件記事をこれらの要素をもつ犯罪スキーマとして定義した．

人物スロットは，ロールスロット，プロフィールスロット，行動スロット．プロフィールススロットを持つ．ロットは，名前スロット，年齢スロット，職業スロット，住所スロットを持つ．

実装されているスロットについて評価を行った．表 1 に評価結果を示す．この表からみても分かるように概ね 8 割程度の正解率が得られている．今回，目的とすることは，犯罪スキーマを用いて，事件記事を構造化することにあるため，抽出精度は，現段階では十分であると考えている．

表 1: 30 記事に対する抽出結果

	全出現数	システム	正解率		全出現数	システム	正解率
ロール	51	50	98.0%	名前	64	51	79.7%
年齢	64	51	79.7%	住所	44	50	88.0%
職業	35	44	79.5%	罪状	30	25	83.3%
供述	7	7	100.0%	動機	30	20	66.7%

3.6.2 新聞の事件記事における時間的照応解析

物語のような文章をコンピュータに理解させるには文 (動作) の時間構造を正確に把握させる必要がある．時間構造記述の方法として時間に関する照応 (時間照応) があり，これを理解させることで文章の深い理解が可能になる．時間照応とは，一般的な名詞句に関する照応関係を時間情報に関して述べたもので，時間情報そのものを指したり，動作同士の時間順序関係を表すものである．時間照応の例を以下に示す．

「買い物に行った．その後夕食を作った.」この例では「行った」と「その後」が照応関係にあり,「行った」と「作った」の時間順序関係を示している．

名詞に関する照応関係と違うところは，先行詞として名詞句だけではなく，文が記述する動作 (動詞句) を指すことがあるという点である．時間照応の先行詞として,「時間表現 (時点・時区間)」,「動作を示すもの (動詞・サ変名詞)」の 2 種類に分類される．照応詞に関しては「時間表現のみを指す (その日など)」,「動作のみを指す (食べた後など)」,「両方指すことができる (その後など)」の 3 種類に分類される．

照応解析に関しては文章内前方照応のみを対象とする．解析の手順として，まず，照応詞を判定し，次に，先行詞候補の探索，先行詞候補と照応詞の照応性の判定の順で行う．

また，解析アルゴリズムとして，分類した照応関係ごとに，以下のように行う．照応性判定に用いる要素は,

1. 指示詞による時制の制限：各指示詞により指せる時制が異なるのを利用して照応性の判定をする．
2. 先行詞候補と照応詞の時間の次元：先行詞候補と照応詞の時間の次元が違うならその照応性を認めない．
3. 照応詞と先行詞のアスペクト性：照応詞が時区間のみを指せるものの場合 (その間など)，先行詞はアスペクト性を持つ動詞，継続相の動詞，もしくは時区間を示す語に限定する．
4. 動詞の格の類似性：同じ動作を指している場合，格が異なっているとその照応性を認めない

なお，本報告書執筆時点で本部分課題の検討が進行中であり，実装，評価は今後の課題である．

3.7 論説文の論理構造解析

3.7.1 語の分布エントロピーと文間の関係解析による論説文書の依存構造解析

我々の方法は，大域的な解析と隣接する文間の解析から成る．大域的な解析では，語が含まれているかどうかについての情報量 (エントロピー) が最小になるようなセグメントの分割を求めることを基本原理とし，文間の解析では，主題の扱いに関する結束性に着目する．

3 分割のモデルで検討する．セグメント (S) に語 w が含まれる生起確率は $\alpha = freq(w, S)/|S|$ であるので，w が S に含まれるかどうかを伝えるメッセージの情報量 (エントロピー) は，

$$H(w, S) = \alpha \log \frac{1}{\alpha} + \bar{\alpha} \log \frac{1}{\bar{\alpha}}$$

となる．ただし，$\bar{\alpha} = 1 - \alpha$ である．セグメント S が S_1，S_2，S_3 に分割されたとき，平均のエントロピーを得られる．

$$H(w, \{S_1, S_2, S_3\}) = \sum_i \frac{|S_i|}{|S|} H(w, S_i)$$

3 分割のモデル (S_1，S_2，S_3) における語 w についての最良のセグメント分割は，

$$\underset{S_1, S_2, S_3}{\operatorname{argmin}} H(w, \{S_1, S_2, S_3\})$$

となる．

前節のクラスタリングによりある程度大きなクラスタを構成する語のグループは，他の語と独立してセグメントを構成する．このようなセグメントは，セグメントどおしあまり重ならないものと予想され，これによりテキスト全体のセグメンテーションを行う．

以上のエントロピーを基にして得られるセグメントの境界は大域的な分割としては合理的であるものの絶対的なものではなく，文と文の隣接関係などの微妙な関係については，主題構造 (前述) に着目してセグメンテーションを微調整する．

3.7.2 論説文における論理構造の解析と視覚化インタフェースの提案

文章構造の解析結果は,「論理構成」という観点から文章評価へと応用できる．この研究では，論説文における話題の移り変わりに着目した筆者の論理の進め方についてモデル化し，それに基づいた文章解析を行う．また，応用として解析結果を視覚化して表示し，人間が文章評価を行う際の支援となるようなユーザインタフェースを提案し実現した．

3.8 本研究のまとめと今後の展望

本課題では，情報検索結果の知的提示における基幹技術として自動要約を位置付け，その手法ならびに利用者インタフェースに関する研究を行なった．今後の展望ならびに課題は下記のとおりである．

- 情報要求指向の単文書要約と複数文書要約をつなぎ目なく接続する枠組の検討
- 複数文書要約に対応する使い勝手の良い利用者インタフェースの構築
- 新聞の事件記事における時間的照応解析についての実装ならびに評価

参考文献

[CG98] Jaime Carbonell and Jade Goldstein. The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries. In *Proceedings of the 21st Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, pp. 335–336, 1998.

[CKT92] Douglass R. Cutting, David R. Karger, and Jan O. Pedersen John W. Tukey. Scatter/gather: A cluster-based approach to browsing large document collections. In *Proceedings of SIGIR '92: 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 318–329, 1992.

[HSI01] Tsutomu Hirao, Yutaka Sasaki, and Hideki Isozaki. An extrinsic evaluation for question-biased text summarization on qa tasks. In *Proceedings of the NAACL 2001 workshop on Automatic Summarization*, pp. 61–68, 2001.

[TS98] A. Tombros and M. Sanderson. Advantages of Query Biased Summaries in Information Retrieval. In *Proceedings of the 21st Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, pp. 2–10, 1998.

4 研究発表

学会誌等

1. 森 辰則，國分 智晴，田中 崇: “空間分割型 CL-LSI による大規模言語横断情報検索”, 情報処理学会論文誌:データベース, Vol. 43, No. SIG 2(TOD 13), pp.27–36, 2002.
2. 森 辰則: “検索結果表示向け文書要約における情報利得比に基づく語の重要度計算”, 自然言語処理, Vol. 9, No. 4, pp.3–32, 2002.
3. 中川 裕志，湯本 紘彰，森 辰則: “出現頻度と連接頻度に基づく専門用語抽出”, 自然言語処理, Vol. 10, No. 1, pp.27–45, 2003.
4. Hiroshi Nakagawa and Tatsunori Mori: “Automatic Term Recognition based on Statistics of Compound Nouns and their Components”, Terminology, Vol. 9, No. 2, pp.201-219, 2003.
5. Tatsunori Mori, Tomoharu Ohta, Katsuyuki Fujihata and Ryutaro Kumon: “An A* Search in Sentential Matching for Question Answering”, IEICE Transactions on Information and Systems, Vol. E86-D, No. 9, pp.1658–1668, 2003.
6. 鈴木朋子, 田村直良: “音声物理量からの抑うつ傾向判定”, 電子情報通信学会論文誌, Vol. J87-D-II, No. 6, pp.1349-1358, 2004.
7. Tatsunori Mori, Masanori Nozawa and Yoshiaki Asada: “Multi-Answer-Focused Multi-Document Summarization Using a Question-Answering Engine”, ACM Transactions on Asian Language Information Processing (TALIP), to appear.
8. Tatsunori Mori: “Japanese Question-answering System Using A* Search and Its Improvement”, ACM Transactions on Asian Language Information Processing (TALIP), to appear.
9. Tatsunori Mori: “Re-interpretation of Rules for Named Entity Task by Probability Assignment”, Natural Language Processing Pacific Rim Symposium ’01 (NLPRS’01), pp.221–228, 2001.
10. Tatsunori Mori: “Segmented LSI for Fully Automated Large-scale Cross-Language Information Retrieval”, Natural Language Processing Pacific Rim Symposium ’01 (NLPRS’01), pp.385–392, 2001.
11. Tatsunori Mori: “Information Gain Ratio as Term Weight — The case of Summarization of IR Results ”, The 19th International Conference on Computational Linguistics (COLING 02), pp.688–694, 2002.
12. Hiroshi Nakagawa and Tatsunori Mori: “Simple but Powerful Automatic Term Extraction Method”, The second International Workshop on Computational Terminology (COMPTERM 02), pp.29–35, 2002.
13. Tatsunori Mori, Masanori Nozawa and Yoshiaki Asada: “Multi-Answer-Focused Multi-Document Summarization Using a Question-Answering Engine”, The 20th International Conference on Computational Linguistics (COLING 04), pp.439–445, 2004.

口頭発表等

1. 浅野秀胤，田村直良: “文章セグメントの単一化による多文章自動要約”, 言語処理学会第 8 回年次大会, pp.551–554, 2002.
2. 大村高史，田村直良: “主題構造解析による新聞記事からの気象情報の抽出と応用”, 言語処理学会第 8 回年次大会, pp.623–626, 2002.
3. 吉田和史，塩田好伸，森辰則: “情報利得比に基づく重要語抽出による情報ナビゲーション”, 言語処理学会第 8 回年次大会,pp.475–478, 2002.
4. Tatsunori Mori, Tomohiro Ohta, Katsuyuki Fujihata and Ryutaro Kumon: “A* Search Algorithm for Question Answering”, Proceedings of NTCIR Workshop 3 Meeting — Part IV: question Answering Challenge (QAC1), pp.39–46, 2002.
5. 金山 淳一，北條 孝，田村 直良: “文章の構造解析による新聞記事からの事件情報抽出”, 情報処理学会自然言語処理研究会 2002-NL-152,pp.1–6, 2002.
6. Tatsunori Mori and Takuro Sasaki: “Information Gain Ratio meets Maximal Marginal Relevance — A method of Summarization for Multiple Documents ”, —Proceedings of NTCIR Workshop 3 Meeting — Part V: Text Summarization Challenge 2 (TSC2), pp.25–32, 2002.
7. 金山淳一, 田村直良: “犯罪スキーマに基づく新聞記事からの事件情報抽出”, 言語処理学会第 9 回年次大会,pp.73–76, 2003.
8. 北條孝, 田村直良: “語の分布と主題の展開に着目した文章の構造化”, 言語処理学会第 9 回年次大会,pp.605–608, 2003.
9. 春日隆緒, 田村直良: “文章のセグメント間関係解析に基づく文章構造解析”, 情報処理学会 自然言語処理研究会 2003-NL-155, pp.59–64, 2003.
10. 伊藤直之, Nikolay Elenkov, 森辰則: “情報検索ナビゲーションにおけるユーザへの提案機構と可視化インタフェース”, 言語処理学会第 10 回年次大会, 2004.
11. Tatsunori Mori, Masanori Nozawa, and Yoshiaki Asada: “Multi-Document Summarization Using a Question-Answering Engine: Yokohama National University at TSC3”, Working Notes of the Fourth NTCIR Workshop Meeting, 2004.
12. Tatsunori Mori: “Japanese Q/A System using A* Search and Its Improvement: Yokohama National University at QAC2”, Working Notes of the Fourth NTCIR Workshop Meeting, 2004.
13. 川本 大輔, 田村 直良: “論説文における論証構造解析と視覚化インターフェィスの提案”, 言語処理学会第 11 回年次大会, 2005.
14. 川口晋平, 森辰則: “一問一答型質問応答を利用した関連質問群に対する質問応答”, 自然言語処理研究会報告 2005-NL-169, 情報処理学会, 2005.
15. Tatsunori Mori and Shinpei Kawaguchi: “Answering Contextual Questions Based on the Cohesion with the Knowledge : Yokohama National University at NTCIR-5 QAC3”, In Proceedings of the Fifth NTCIR Workshop Meeting, 2005.

特許出願・取得

1. 国立大学法人横浜国立大学: “文書要約システム及び文書要約方法及びプログラムを記録したコンピュータ読み取り可能な記憶媒体及びプログラム”, 特願 2004-239544, 未公開.

公開ソフトウェア

1. 中川 裕志，森 辰則: 専門用語抽出システム

   URL: http://www.forest.eis.ynu.ac.jp/TermExtraction/

   与えられた特定分野のコーパスのみから，その中に現れる専門用語を特定し，抽出することができる.